Structure
Article

Structures of Human Pumilio with Noncognate RNAs Reveal Molecular Mechanisms for Binding Promiscuity

Yogesh K. Gupta,[1] Deepak T. Nair,[1] Robin P. Wharton,[2] and Aneel K. Aggarwal[1,*]
[1]Department of Structural and Chemical Biology, Mount Sinai School of Medicine, Box 1677, 1425 Madison Avenue, New York, NY 10029, USA
[2]Howard Hughes Medical Institute, Department of Molecular Genetics, & Microbiology and Department of Cell Biology, Box 3657, Duke University Medical Center, Durham, NC 27710, USA
*Correspondence: aneel.aggarwal@mssm.edu
DOI 10.1016/j.str.2008.01.006

SUMMARY

Pumilio is a founder member of the evolutionarily conserved Puf family of RNA-binding proteins that control a number of physiological processes in eukaryotes. A structure of human Pumilio (hPum) Puf domain bound to a *Drosophila* regulatory sequence showed that each Puf repeat recognizes a single nucleotide. Puf domains in general bind promiscuously to a large set of degenerate sequences, but the structural basis for this promiscuity has been unclear. Here, we describe the structures of hPum Puf domain complexed to two noncognate RNAs, $CycB_{reverse}$ and Puf5. In each complex, one of the nucleotides is ejected from the binding surface, in effect, acting as a "spacer." The complexes also reveal the plasticity of several Puf repeats, which recognize noncanonical nucleotides. Together, these complexes provide a molecular basis for recognition of degenerate binding sites, which significantly increases the number of mRNAs targeted for regulation by Puf proteins in vivo.

INTRODUCTION

Pumilio (Pum) was first identified as a factor required for abdominal patterning in early *Drosophila* embryos (Lehmann and Nusslein-Volhard, 1987). Pum was later shown to regulate abdominal patterning by binding to a pair of 32 nucleotide Nanos Response Elements (NREs) within the 3′-untranslated regions (3′UTR) of hunchback (hb) mRNA and repressing its translation (Murata and Wharton, 1995). Each NRE is bipartite and contains Box A (GUUGU) and Box B (AUUGUA) sequences (Wharton and Struhl, 1991; Zamore et al., 1997); the binding of Pum to these Boxes provides a platform for recruitment of the essential *trans*-acting cofactors, Nanos (Nos) and Brain Tumor (Brat) (Sonoda and Wharton, 1999, 2001). The mechanism by which the resulting Pum/Nos/Brat/NRE quaternary complex blocks mRNA translation is unclear but is thought to involve both poly(A)-tail-dependent and poly(A)-tail-independent regulation (Chagnovich and Lehmann, 2001; Cho et al., 2006; Wharton and Struhl, 1991; Wreden et al., 1997).

Pum and Nos are thought to act together to regulate other processes including learning and memory (Dubnau et al., 2003; Menon et al., 2004; Ye et al., 2004) as well as many facets of primordial germ cell (PGC) biology (Asaoka-Taguchi et al., 1999; Lin and Spradling, 1997). The relevant mRNA targets for the former processes have not yet been definitively identified (Wharton and Aggarwal, 2006). But two mRNAs that are likely direct targets of Pum and Nos in the PGCs have been identified: head involution defective (hid) and maternal CyclinB (CycB) mRNAs (Kadyrova et al., 2007; Sato et al., 2007). Maternal CycB mRNA bears a Nanos Respsonse Element (NRE) that binds Pum and Nos in its 3′UTR (Kadyrova et al., 2007). Mutations in either of the two UGUA motifs within the NRE abolish Pum binding in vitro and translational regulation in vivo (Kadyrova et al., 2007).

Pum is a founder member of the novel class of RNA-binding proteins (Wharton et al., 1998; Zamore et al., 1997). The similarity between the RNA-binding domains (RBDs) of Pum and the *C. elegans* translational repressor FBF (Zhang et al., 1997) defined a Puf (*Pu*m and *F*BF) domain, which is conserved in organisms as diverse as plants, yeast, worms, and humans. All Puf proteins examined to date bind to elements in the 3′UTRs of specific mRNAs and thereby repress translation and/or stimulate decay (Wickens et al., 2002). FBF, for example, promotes mitosis and a switch from spermatogenesis to oogenesis by repressing the translation of gld-1 and fem-3 mRNAs, respectively (Bernstein et al., 2005; Crittenden et al., 2002; Zhang et al., 1997). The yeast proteins Puf3, Puf4/Puf5, and Puf6 regulate expression of Cox17, HO, and Ash1 mRNAs, respectively (Gu et al., 2004; Hook et al., 2007; Jackson et al., 2004; Olivas and Parker, 2000; Tadauchi et al., 2001). The Puf domain is characterized by eight imperfect repeats of ~36 amino acids (Puf repeats), followed by a C-terminal extension. Crystal structures of fly and human Puf domains have revealed an extended, rainbow-shaped molecule (Edwards et al., 2001; Wang et al., 2001), composed of tri-α-helical Puf repeats packed in tandem. Crystal structures of human Pumilio (hPum) Puf domain in complex with fragments of hb NREs showed modular binding of each Puf repeat to a single RNA base, with contacts mediated primarily by three amino acids at conserved positions of each repeat (Wang et al., 2002).

Biochemical experiments on Puf domain proteins have posed a curious dichotomy. The modular "one nucleotide-one Puf repeat" recognition pattern has been exploited to engineer new binding specificities (Cheong and Hall, 2006; Wang et al.,

2002), suggesting that relatively simple rules for sequence recognition by Puf domains might be elucidated. However, despite a strong conservation of residues that contact RNA bases in the hPum-hb NRE structures, Puf proteins bind to a variety of sites in vivo that differ in sequence. A detailed study of the hb and CycB NREs bound by Pum, for example, reveals significant variation in sequence outside the core "UGU" triplet (Kadyrova et al., 2007; Murata and Wharton, 1995). Similarly, gld-1 and fem-3 elements bound by FBF, and Cox17, HO, and Ash1 elements bound by Pufs 3-6, differ in sequence outside of the UGU triplet (Bernstein et al., 2005; Crittenden et al., 2002; Gu et al., 2004; Hook et al., 2007; Jackson et al., 2004; Olivas and Parker, 2000; Tadauchi et al., 2001; Zhang et al., 1997). In addition, a genome-wide analysis of Pum-bound RNAs reveals a significant fraction (~20%) that do not even bear a UGU triplet (Gerber et al., 2006), and a microarray analysis has identified hundreds of potential RNA targets for Puf1–Puf5 in yeast (Gerber et al., 2004). Strikingly, Puf4 and Puf5 bind preferentially to sites longer than the canonical 8 nucleotides (nts) (Gerber et al., 2004), and FBF has also been shown to require the presence of an extra nucleotide in the middle of its recognition site (Opperman et al., 2005). These and other recent studies posit a more complex picture of Puf-RNA binding than that gleaned initially from the hPum-hb NRE structures. That is, many of the sites have radically different sequences and some are longer than the canonical 8 nts. How Puf proteins tolerate such variation in RNA sequence and length is not well understood.

To better understand the structural basis for the binding promiscuity of Puf domains, we have solved the structures of hPum Puf domain bound to two noncognate RNAs, $CycB_{reverse}$ and Puf5. We show that Puf proteins can exhibit broader specificity in two ways: ejection of nucleotides from their stacked arrangement (allowing recognition of longer sites) and recognition of multiple nucleotides by a given Puf repeat.

RESULTS

hPum-$CycB_{reverse}$ Complex

Because of our interest in the role of Pumilio in the regulation of CycB in PGCs, we prepared for binding experiments a fluorescein (Fl)-labeled RNA fragment ($CycB_{direct}$), 5′-Fl-UUUGUAA UUU-3′, which encompassed the first "UGUA" motif (or Box 1) of the CycB NRE (Figure 1A). As a control, we also prepared a noncognate RNA fragment ($CycB_{reverse}$), 5′-Fl-UUUAAUG UUU-3′, with the sequence running in the reverse direction. Surprisingly, when we measured the binding of hPum to these two RNAs by fluorescence anisotropy (Figure 1C), hPum Puf domain bound to $CycB_{reverse}$ with only ~7-fold lower affinity (K_d of 23.9 nM) than to $CycB_{direct}$ (K_d of 3.3 nM). To understand the basis for this relatively efficient recognition of the noncanonical reversed sequence, we cocrystallized hPum Puf domain with $CycB_{reverse}$, obtaining cocrystals that diffracted to 2.5 Å resolution, and which contained two complexes (A and B) in the asymmetric unit. The structure was determined by molecular replacement by using the protein from the hPum-hb NRE complex as a search model (Wang et al., 2002), and a bromine derivative was used to confirm the register of the RNA. Although, the bromine derivative (2.8 Å) diffracted to lower resolution than the native cocrystals, the RNA densities were slightly better defined, and it was therefore used for further refinement. Also, the RNA densities were much better defined in one of the two complexes in the asymmetric unit, and thus only the RNA chain in this complex (A) was built. The final model contains hPum A (amino acids 828–1168) bound to $CycB_{reverse}$ A (nucleotides 1–9), hPum B (amino acids 829–1168), and 145 solvent molecules.

From the structure, $CycB_{reverse}$ binds to the inner concave surface of a curved hPum molecule (Figure 2). The binding is in the normal forward direction with the 5′ and 3′ ends of the RNA near the C and N termini of the protein, respectively (Wang et al., 2002). As in previous structures (Edwards et al., 2001; Wang et al., 2001, 2002), hPum is composed of eight tandem tri-α-helical Puf repeats, wherein one set of helices cover its outer convex surface, another set forms the ridge of the molecule, and another set coats the inner concave surface (Figure 2). It is amino acids from these "inner" helices that interact directly with the RNA. More specifically, as in hb NRE complexes (Wang et al., 2002), hydrophilic and/or charged amino acids at positions 12 (Asn/Ser/Cys) and 16 (Gln/Glu) along these repeats make direct contacts with the edges of the bases, while an amino acid at position 13 (Tyr/His/Arg/Asn) stacks in between the bases (Figures 1B, 2, and 3). In addition, amino acids at position 10 (His/Phe/Tyr/Asn) present a "wall" of van der Waals contacts that constrain the positions of the sugars of the RNA chain. However, unlike the hb NRE complexes, the Puf repeats do not recognize successive nucleotides. That is, U_6 (5′-$U_1U_2U_3A_4A_5\underline{U_6}G_7U_8U_9U_{10}$-3′) acts as a "spacer" and is excluded from the hPum binding surface (Figure 2). Four of the eight repeats contact nucleotides that are different from those in the hPum-hb NRE complexes (Wang et al., 2002).

The entire U_6 nucleotide is displaced from the Pum binding surface, and the base is completely solvent exposed (Figures 2 and 4). Nonetheless, the nucleotides on either side ($\underline{A_5}$ and G_7) of U_6 are relatively unperturbed in structure and maintain their "continuity" against the Pum binding surface. The most notable change is an ~4 Å shift in the A_5 sugar, which permits new van der Waals contacts with the tyrosine at position 10 on repeat 5. This same tyrosine along with glutamine at position 9 on repeat 4 also facilitate eviction of U_6 by making van der Waals contacts with the displaced phosphate and sugar, respectively (Figures 3 and 4). In all, the expulsion of U_6 permits G_7, U_8, and U_9 to correctly align against repeats 3, 2, and 1, respectively.

Interactions that are different from those in hb NRE complexes are made by repeats 7, 4, 3, and 1 (Figure 4). They recognize U_2, A_5, G_7, and U_9 (5′-$U_1\underline{U_2}U_3A_4\underline{A_5}U_6\underline{G_7}U_8\underline{U_9}U_{10}$-3′) as compared to G_2, U/C_5, A_6, and A_8 (5′-A-$_2$U-$_1U_1\underline{G_2}U_3A_4\underline{U/C_5}\underline{A_6}U_7\underline{A_8}$-3′) in the hb NRE complexes (Wang et al., 2002). The most interesting of these noncanonical interactions is that of repeat 7 (bearing serine at position 12 and glutamate at position 16) with U_2. These residues of repeat 7 have been thought to be highly selective for recognition of the conserved G of the UGU triplet that lies at the core of most Puf binding sites, and thus the accommodation of U_2 by this repeat is surprising. However, although the U_2 sugar is fixed by a hydrogen bond to Tyr1123, the electron density for the base is not well defined. This suggests certain mobility when noncognate U_2 is bound by repeat 7 and, as currently built, there is the possibility of only a single hydrogen bond between N3 of U_2 and the glutamate at position 16 (Figures 2–4). This smaller number of interactions with U compared to G may

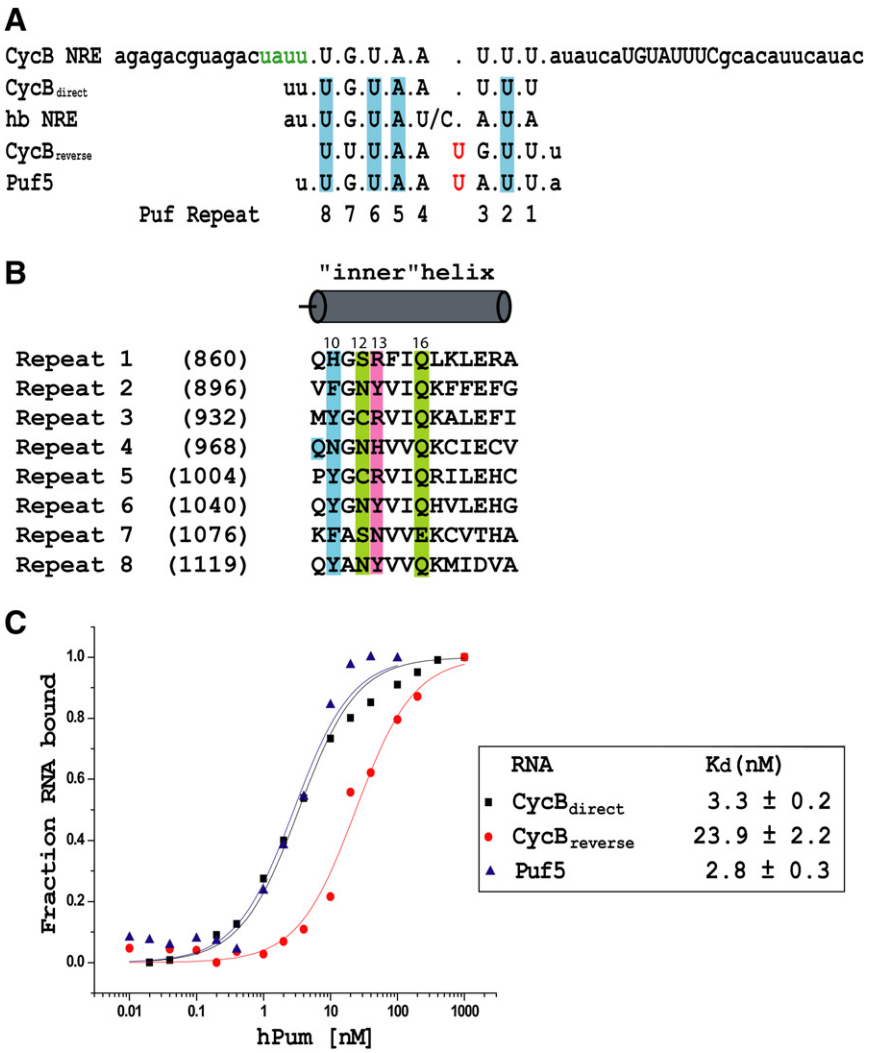

Figure 1. Cognate and Noncognate RNAs

(A) Alignment of cognate ($CycB_{direct}$ and hb NRE) and noncognate ($CycB_{reverse}$ and Puf5) RNA sequences against the full CycB NRE. The two Pum binding sites in CycB NRE are shown in upper case (Box 1 at 5′ end and Box 2 at 3′ end). $CycB_{direct}$ contains the Box 1 sequence; $CycB_{reverse}$ contains the Box 1 sequence in reverse (running 3′-5′ instead of 5′-3′); hb NRE contains the Box B sequence of full hb NRE; and Puf5 contains the sequence identified as a binding site for the yeast Puf5 protein. Highlighted in blue are the nucleotides conserved in hb-NRE, $CycB_{reverse}$, and Puf5 structures, against PUF repeats 8, 6, 5, and 2. Highlighted in red is the flipped-out nucleotide, U_6, in $CycB_{reverse}$, and Puf5 structures.

(B) Alignment of residues on the "inner" helix of the eight hPum Puf repeats. Residues at position 10, 12, 13, and 16 are highlighted.

(C) Binding curves for $CycB_{direct}$, $CycB_{reverse}$, and Puf5 RNAs.

account for the decrease in binding when G is substituted by U at the cognate position in the hb NRE (Cheong and Hall, 2006). Repeat 4 recognizes A_5, as compared to U_5 or C_5 in the hb NRE complexes, with a hydrogen bond from glutamine to N1 of A_5 replacing a bond to O2 of U_5/C_5 in the hb NRE structures. In addition, C2 of A_5 makes van der Waals contacts with an asparagine, the equivalent of which are not seen with U_5/C_5 in the hb NRE complexes (Figures 3 and 4). These additional van der Waals interactions may provide a basis for the preference of an A at this position (see Discussion). Repeat 3 recognizes G_7 instead of A_6 in the hb NRE complexes. This novel pairing is specified by hydrogen bonds with both glutamine and cysteine on repeat 3. Repeat 1 binds U_9 with the glutamine making hydrogen bonds with both the N3 and O2 atoms of the base. Interestingly, Arg864, which stacks with A_8 in the hb NRE structures, points away and Tyr900 (from repeat 2) makes a hydrogen bond with the O4′ of U_9 sugar. In all, the noncognate nucleotides A_5, G_7, and U_9 are accommodated surprisingly well opposite repeats 4, 3, and 1, respectively, while U_2 opposite repeat 7 is bound much less well (Figures 3 and 4). These observations agree with binding studies of hPum to the hb NRE, which show relatively mild effects on affinities when cognate nucleotides opposite repeats 3 and 1 are mutated, but a much more marked effect when G opposite repeat 7 is mutated (Cheong and Hall, 2006).

Interactions in common with hb NREs are limited to repeats 8, 6, 5, and 2. They recognize U_1, U_3, A_4, and U_8 (5′-$\underline{U}_1U_2\underline{U}_3\underline{A}_4A_5U_6G_7\underline{U}_8U_9U_{10}$-3′), as compared to U_1, U_3, A_4, and U_7 (5′-$A_{-2}U_{-1}\underline{U}_1G_2\underline{U}_3\underline{A}_4U_5A_6\underline{U}_7A_8$-3′) in the hb NRE complexes. More specifically, as in the hb NRE complexes, repeats 8, 6, and 2 recognize Us via an asparagine and a glutamine that make hydrogen bonds with the polar atoms (O4, N4, and O2) at the Watson-Crick edge of the base, while a tyrosine stacks on the 3′ side of the base (Figures 2 and 3). Repeat 5 recognizes A_4 via a hydrogen bond between a glutamine and the N1 atom of the base and van der Waals contacts between a cysteine and the C2 atom of the base.

hPum-Puf5 Complex

The hPum-$CycB_{reverse}$ complex suggested similarities between $CycB_{reverse}$ and another noncognate RNA, the Puf5 sequence, 5′-UUGUAAUAUUA-3′, which has been identified as a binding site for the yeast Puf5 protein in a genome-wide microarray analysis (Figure 1A) (Gerber et al., 2004). In particular, if one allows for a "spacer" nucleotide, seven of ten nucleotides align between $CycB_{reverse}$ and Puf5. Thus, we were intrigued as to whether hPum would also bind Puf5 with a flipped-out nucleotide. We obtained hPum-Puf5 cocrystals that diffracted to 2.3 Å resolution, and the structure was determined by molecular replacement. A bromine derivative was used to confirm the register of the RNA. The RNA densities were very well defined in both complexes of the asymmetric unit. The final model contains hPum A (amino acids 828–1168) bound to Puf5 A (nucleotides 1–10), hPum B bound to Puf5 B (nucleotides 2–10), and 246 solvent molecules.

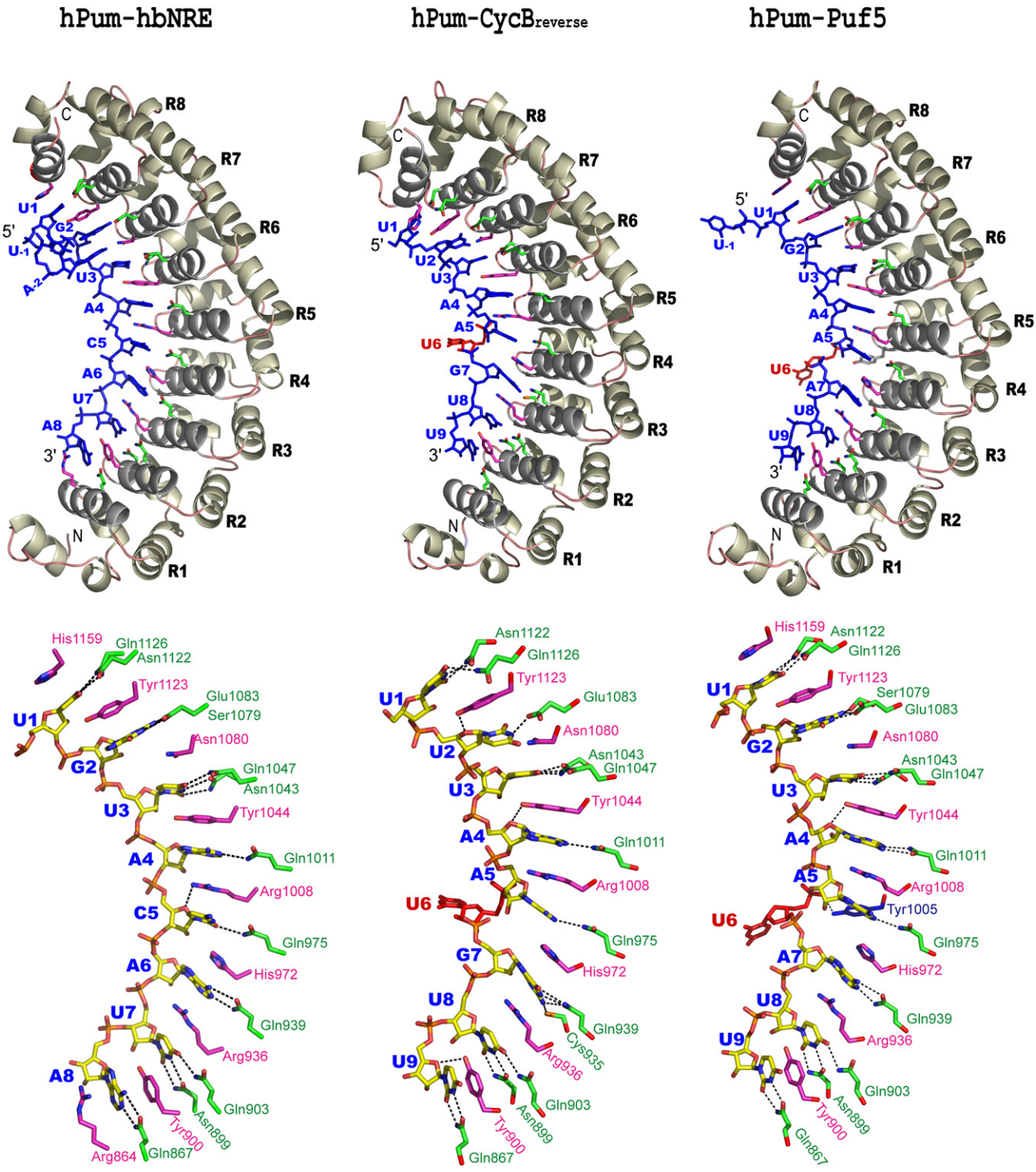

Figure 2. Structures of hPum-RNA Complexes

The upper panel shows structures of hPum bound to hb-NRE (left), $CycB_{reverse}$ (middle), and Puf5 (right) RNAs. hPum is shown in ribbon, and the RNAs are depicted in stick mode. RNA bases are aligned along the concave surface of the protein; the flipped-out U_6 nucleotide in $CycB_{reverse}$ and Puf5 structures is highlighted in red. The inner helices are shown in a darker gray color than the outer helices. Residues that make hydrogen bonds and stacking interactions with the bases are shown in green and magenta, respectively. The lower panel shows the protein-RNA interface in more detail. As above, residues that make hydrogen bonds and stacking interactions with bases are shown in green and magenta, respectively. The flipped-out U_6 nucleotide in $CycB_{reverse}$ and Puf5 structures is highlighted in red. For clarity, U_{-1} and A_{-2} in hb-NRE and U_{-1} in Puf5 are omitted.

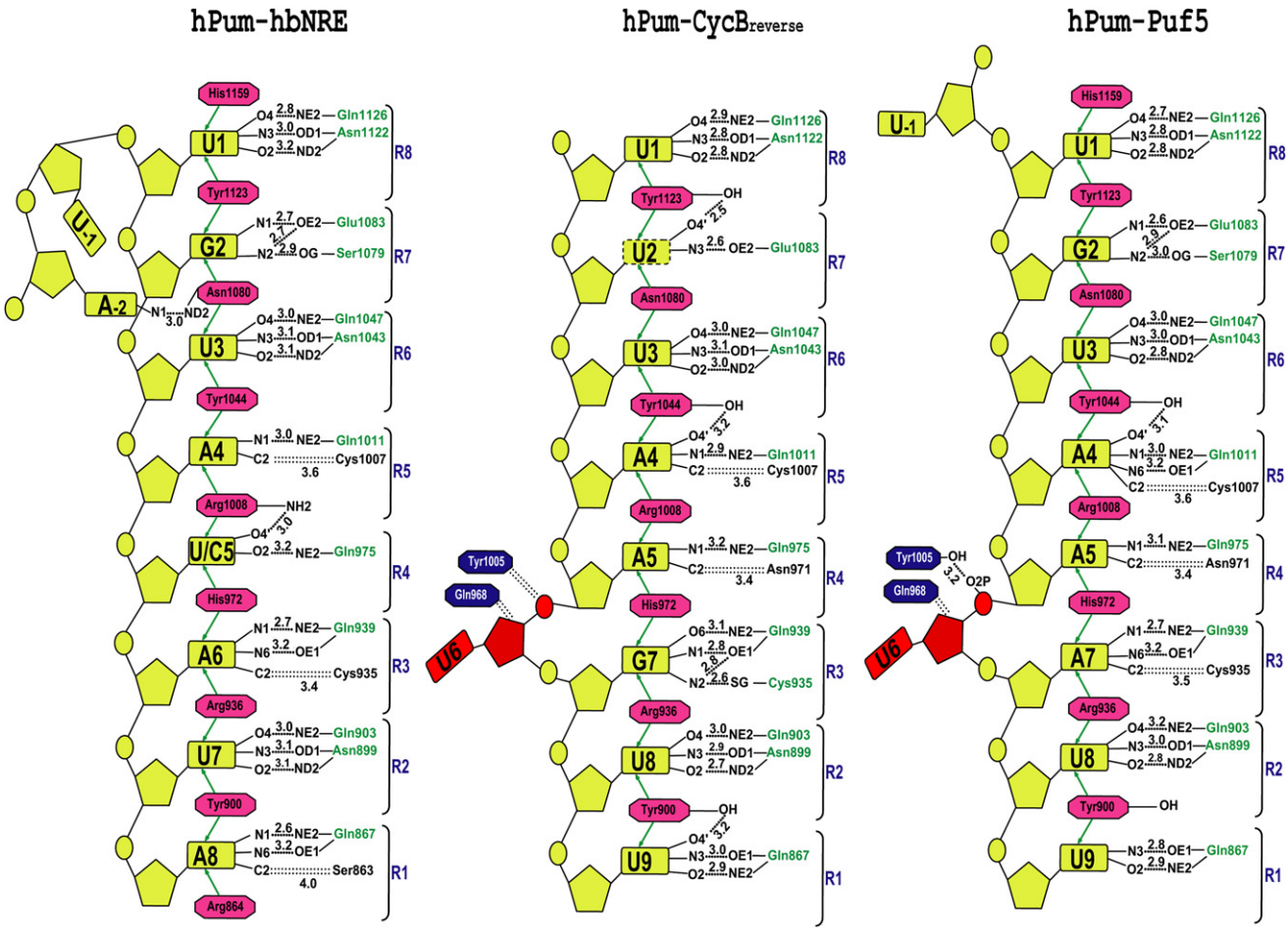

Figure 3. A Schematic of hPum-RNA Interactions

Schematic representations of hPum-RNA interactions in hb-NRE, $CycB_{reverse}$, and Puf5 structures. Hydrogen bonds are depicted as continuous single black dots (……) and distances are shown on the top of the dots. van der Waals contacts are shown by double dotted lines (:::::::), and stacking interactions by green arrows. The U_2 base in $CycB_{reverse}$ structure is shown in a dotted box because of its weak density and tentative contacts.

As with $CycB_{reverse}$, hPum does not interact with successive nucleotides on the Puf5 sequence (5′-U-$_1$$U_1$$G_2$$U_3$$A_4$$A_5$$U_6$$A_7$$U_8$$U_9$$A_{10}$-3′). The same relative nucleotide, U_6, is expelled from the hPum binding surface as that observed with $CycB_{reverse}$ (Figure 2). Interestingly, hPum binds Puf5 with a K_d (2.8 nM) that is similar to that for the cognate $CycB_{direct}$ sequence (3.3 nM) (Figure 1C). hPum's high affinity for Puf5 (compared to $CycB_{reverse}$) likely derives from the presence of a canonical UGUA motif. These four nucleotides are recognized in the same way as in hb NREs (Wang et al., 2002), where repeats 8 and 6 recognize U_1 and U_3, repeat 7 recognizes G_2, and repeat 5 recognizes A_4 (Figures 2 and 3). The hydrogen bonding pattern (and van der Waals contacts) is almost identical to that observed in the hb NRE complexes. The next two nucleotides, A_5 and U_6, are analogous to the nucleotides in $CycB_{reverse}$. That is, A_5 makes a hydrogen bond and van der Waals contacts with repeat 4, while U_6 is ejected from the binding surface. The exclusion of U_6, in this case, allows A_7, U_8, and U_9 to dock against repeats 3, 2, and 1, respectively (Figures 2 and 3). A_7 and U_8 are analogous to A_6 and U_7 in hb NREs, while U_9 is analogous to U_9 in $CycB_{reverse}$; and they partake in the same combination of hydrogen bonds and van der Waals contacts as those described above for hb NREs and $CycB_{reverse}$. In all, hPum recognizes Puf5 by the mix of modes used to recognize hb NREs and $CycB_{reverse}$. Out of the eight nucleotides that interact with repeats, six are recognized in an hb NRE-like manner, and two in a $CycB_{reverse}$-like manner (Figure 4). The alignment of these nucleotides is determined by the exclusion of U_6. As with $CycB_{reverse}$, the "external" position of U_6 is stabilized in part by contacts with a tyrosine on repeat 5 and a glutamine on repeat 4 (Figure 4). Thus, these two amino acids take on added significance in the Puf5 and $CycB_{reverse}$ structures.

DISCUSSION

The Puf family members control a number of physiological processes in eukaryotes, and they generally bind to a large set of degenerate RNA sequences. The $CycB_{reverse}$ and Puf5 structures reveal two mechanisms for promiscuous binding of Puf proteins to such degenerate or noncognate sequences.

The first mechanism is the ejection of an "undesirable" nucleotide from the RNA binding surface. In both $CycB_{reverse}$ and Puf5

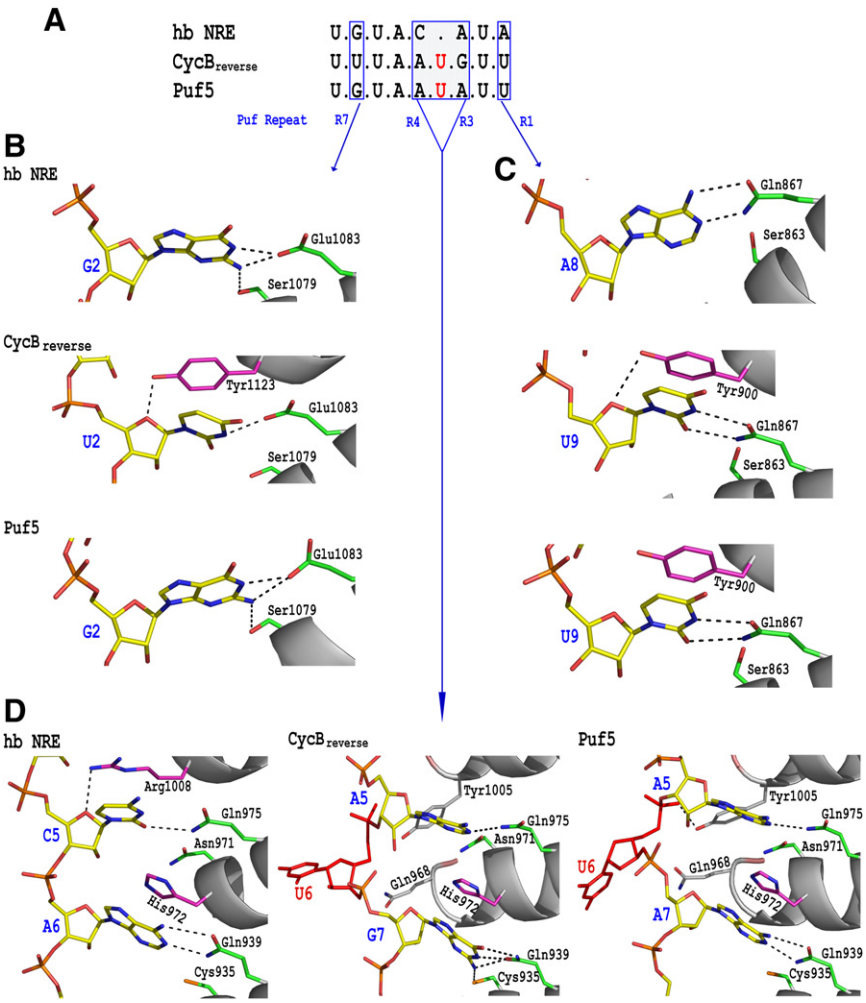

Figure 4. Close-Up Views of Base Interactions

(A) An alignment of hb-NRE, $CycB_{reverse}$, and Puf5 RNA portions that bind the inner concave surface of hPum. Bases that differ are outlined with blue boxes. The flipped-out nucleotide U_6 is highlighted in red.

(B) Recognition of G (hb-NRE and Puf5) or U ($CycB_{reverse}$) by Puf repeat 7.

(C) Recognition of A (hb-NRE) or U ($CycB_{reverse}$ and Puf5) by Puf repeat 1.

(D) Interactions with CA (hb-NRE), AUG ($CycB_{reverse}$), or AUA (Puf5) by Puf repeats 5–3.

structures, U_6 is ejected with relative ease and with little or no disturbance to the bracketing nucleotides. Indeed, amino acids from two adjoining repeats (5 and 4) facilitate eviction of U_6 by making contacts with its sugar and phosphate groups. However, the U_6 base is completely solvent accessible, and it could in principle interact with another protein or RNA in vivo, thereby enhancing specificity. As such, it would add another layer of regulation to translational repression by bringing other protein and/or RNA components to the Puf complex. Interestingly, FBF in *C. elegans* binds its cognate RNA site with a "spacer" nucleotide, but it requires a local distortion in the protein. Specifically, FBF repeat 5 and flanking 16 residues confer the requirement for the spacer nucleotide (Opperman et al., 2005). How these residues locally distort FBF and whether they directly interact with the spacer nucleotide remain to be determined, but they appear to play a more active role in specifying a spacer nucleotide than the ejection of U_6 by hPum. In both $CycB_{reverse}$ and Puf5 structures, it is the same nucleotide (between repeats 4 and 3) that is evicted from the binding surface. However, this is most likely due to the choice of RNA sequences and not because of any structural preference for ejecting a U, or for ejecting a nucleotide along a certain position in the RNA sequence. In all, a bulge of one or more nucleotides can alter the register of a longer RNA sequence and optimize it for Puf binding. Accordingly, Puf4 and Puf5 may bind to their preferred 9 and 10 nt sites by ejecting one and two nucleotides from their binding surfaces, respectively (Gerber et al., 2004). For an optimal register, one can speculate that Puf4 flips the seventh nucleotide (UGUAU/CAA/UUA) and Puf5 flips the sixth and eighth nucleotides (UGUAAC/UAA/UUA); however, the identity of ejected nucleotides may only become clearer through further biochemical analyses and/or crystal structures of the two proteins with their preferred sites.

The second mechanism for promiscuous RNA binding is the plasticity of several Puf repeats in accommodating noncanonical nucleotides. In particular, the identity of the two base interacting residues (at positions 12 and 16) is not sufficient by itself to predict the specificity of each repeat. Thus, although repeat 4 carries an asparagine and a glutamine that are considered the signature recognition residues for a U, we find noncanonical A making optimal hydrogen bonds with this repeat. Indeed, genome-wide analyses of RNAs bound to Pum and yeast Puf3 suggest that an A is actually preferred over U or C at this position (Gerber et al., 2004, 2006). Repeats 3 and 1 also carry residues that are considered selective for an A, but we observe G and U making favorable hydrogen bonds with these two repeats. Consistent with the structure, binding studies on hb NRE with hPum show that noncanonical nucleotides opposite repeats 3 and 1 have relatively mild effects on binding (Cheong and Hall, 2006). The pairing of noncanonical U opposite repeat 1 may turn out be a quite common feature of Puf proteins. FBF, Puf5, and Puf6, for example, appear to bind RNA elements in fem-3, HO and Ash-1 mRNAs, respectively, which position a U opposite repeat 1 (Bernstein et al., 2005; Gu et al., 2004; Hook et al., 2007). The $CycB_{reverse}$ structure shows that even repeat 7, which is considered highly selective for a G, is able to accommodate a noncanonical U opposite it. However, in this case, the density for the base is weak, suggesting that it is both mobile and unable to make proper contacts. The mobility of U may be confounded by the special nature of repeat 7, which carries an asparagine at position 13, the side chain

of which is not long enough to form stacking interactions with a base (Cheong and Hall, 2006).

The plasticity of Puf repeats has a bearing on the engineering of Puf domains with new specificities. In particular, it raises the possibility that residues other than two base interacting residues play an auxiliary role in conferring specificity. Among these "other" residues, the amino acid at position 13 appears to be important. We note that three of the four nucleotides that align between hb NRE, $CycB_{reverse}$, and Puf5 structures are Us opposite repeats 8, 6, and 2, which all carry a tyrosine (at position 13) that stacks on the 3′ side of the base. The stringency of repeats 8, 6, and 2, as compared to repeat 4, could be dictated in part by these stacking interactions with the tyrosines. The length of the tyrosine side chain and its aromatic character lend to much more extensive stacking interactions that appear to "fix" the U for favorable hydrogen bonding.

Overall, the ability of Puf proteins to accommodate noncanonical nucleotides in many ways is a reflection of their modular nature, wherein each Puf repeat recognizes a base more or less independently of the others. The structures presented in this work show that interactions of noncanonical nucleotides with one Puf repeat do not significantly perturb interactions made by neighboring repeats. They also reveal that "undesirable" nucleotides can be ejected from the Puf binding surface, allowing the remaining nucleotides to bind with little or no apparent cost in binding free energy (e.g., interaction of hPum with the $CycB_{direct}$ versus Puf5 sites). The energetics of individual amino acid-nucleotide interactions remain to be fully tested, however. Taken together, the modularity of Puf domains, their ability to bulge nucleotides from the binding surface, and the ability of certain repeats to recognize multiple nucleotides increases the number of sites that Puf proteins can target in vivo and permits the regulation of many more biological processes (Gerber et al., 2004, 2006).

Table 1. Data Collection and Refinement Statistics

	(Br) hPum-$CycB_{reverse}$	hPum-Puf5
Data Collection		
Wavelength(Å)	0.918	0.979
Resolution (Å)	2.8	2.3
R_{sym} (%)[a,b]	7.1 (22.0)	6.4 (16.4)
Number of reflections	17996 (1254)	31513 (2788)
Completeness	92.3 (67.5)	94.0 (84.9)
Redundancy	6.4 (6.3)	3.7 (2.6)
I/sigma	20.3 (4.7)	18.3 (5.2)
Refinement		
Resolution (Å)	50–2.8	50–2.3
R_{factor} (%)[c]/R_{free} (%)[d]	24.4/26.8	21.2/27.7
Nonhydrogen Atoms		
Protein/RNA/Water	5493/184/145	5473/392/246
Average B Factors (Å^2)		
Protein/RNA/Water	44.2/73.8/37.5	39.7/52.3/47
Rms Deviations		
Bond lengths (Å)	0.014	0.016
Bond angles (°)	1.6	1.56
Ramachandran Plot		
Most Favored (%)	90.9	91.5
Additional allowed (%)	8.8	7.5
Generously Allowed (%)	0.3	1.0
Disallowed (%)	0.0	0.0

[a] Values for outermost shells are given in parentheses.
[b] $R_{sym} = \sum|I - < I >|/\sum I$, where I is the integrated intensity of a given reflection.
[c] $R_{factor} = \sum||F_{observed}| - |F_{calculated}||/\sum|F_{observed}|$.
[d] R_{free} was calculated by using 6.7% and 10% of random data omitted from the refinement of hPum-$CycB_{reverse}$ and hPum-Puf5 complexes, respectively.

EXPERIMENTAL PROCEDURES

Preparation of Protein and RNA

The hPum Puf domain (Gly^{828}-Lys^{1172}) was cloned into a derivative of pET19b plasmid as a His-tagged fusion protein. The fusion protein was expressed in *E. coli*, passed over Ni^{2+} column, and then cleaved with TEV protease to remove the His-tag. The protein was further purified by size exclusion chromatography on a Superdex 75 column and then concentrated to 10 mg/ml in a buffer containing 25 mM Tris-Cl (pH 7.4), 0.15 M NaCl, 2 mM DTT, and 5% (v/v) glycerol. Native $CycB_{reverse}$ (5′-UUUAAUGUUU-3′), brominated $CycB_{reverse}$ (5′-UUUAAU^{Br}GU^{Br}UU-3′), native Puf5 (5′-UUGUAAUAUUA-3′), and brominated Puf5 (5′-UUGU^{Br}AAU^{Br}AUU^{Br}A-3′) RNA oligonucleotides were purchased from Dharmacon Research in their 2′-ACE protected form and then deprotected (as recommended by the manufacturer) and resuspended in 10 mM Tris (pH 6.0) and 50 mM NaCl. To remove any secondary structure, the RNAs were heated at 60°C for 10 min and then cooled on ice. Prior to crystallization, hPum was added in equimolar amounts to the RNAs.

Crystallization and Structure Determination

The hPum-$CycB_{reverse}$ complex was crystallized from solutions containing 10% (w/v) PEG 8000, 50 mM sodium cacodylate (pH 6.5), 0.2 M KCl, 0.1 M magnesium acetate (Table 1). The cocrystals belong to space group $P2_12_12_1$ with cell dimensions a = 35.8 Å, b = 65.6 Å, c = 313.9 Å, $\alpha = 90°$, $\beta = 90°$, and $\gamma = 90°$ and with two complexes per asymmetric unit. Cocrystals with brominated $CycB_{reverse}$, native Puf5, and brominated Puf5 RNAs were obtained under similar conditions as above, and they belong to the same space group as the native hPum-$CycB_{reverse}$ cocrystals and have very similar unit cell dimension (hPum-Puf5, for example, has a = 35.8 Å, b = 64.2 Å, c = 321.2 Å). Prior to data collection, the cocrystals were cryo-protected by serial soaks in solutions containing 14%(w/v) PEG 8000, 50 mM sodium cacodylate (pH 6.5), 0.2 M KCl, 0.1 M magnesium acetate and increasing amounts (5%, 10%, 15%, 20%, 25%, 30%) of 2-methyl-2,4-pentanediol (MPD) and then flash frozen in liquid nitrogen.

X-ray data on cryo-cooled native (2.5 Å) and brominated (2.8 Å) hPum-$CycB_{reverse}$ cocrystals were measured at the Advanced Photon Source (APS; beamlines 17-ID and 17-BM). Data on cryo-cooled native (2.32 Å) and brominated (2.5 Å) hPum-Puf-5 cocrystals were measured at Brookhaven National Laboratory (BNL; beamlines X4C and X6A). Data on brominated hPum-$CycB_{reverse}$ and hPum-Puf5 cocrystals were measured at the peak of the Br *K*-edge absorption profile. All data were indexed and integrated with DENZO and reduced with SCALEPACK (Otwinowski and Minor, 1997). The hPum-$CycB_{reverse}$ structure was solved first by molecular replacement (MR) by using only the protein from an hPum-hb NRE complex (PDB code: 1M8Y) as a search model. Specifically, the program PHASER gave a unique MR solution (McCoy et al., 2005), which was then rigid-body refined by using CNS (Brunger et al., 1998). After a first round of positional and B-factor refinement, the R_{factor} dropped to 31%, and the R_{free} was 36%. The resulting maps showed good densities for RNA in one of the two complexes in the asymmetric unit and also indicated that one of the "middle" nucleotides was evicted from the binding surface. Brominated hPum-$CycB_{reverse}$ complex data were used to confirm the register of the RNA and the identity of the evicted nucleotide. Although, the brominated complex diffracted to lower resolution than the native, the RNA densities were slightly better defined with brominated data,

and we used it for further refinement, with the RNA built in only one of the two complexes of the asymmetric unit. After several iterative rounds of refinement with CNS, model building with O (Jones et al., 1991), and water picking, the R_{free} lowered to ~27%. A Ramachandran plot for the refined model, generated with the program PROCHECK (Laskowski et al., 1993), shows 90.9% of the residues in the most favored regions, 8.8% in the additionally allowed regions, 0.3% in the generously allowed regions, and none in the disallowed regions. Residues Arg829, Arg831, Arg837, Lys1138, and Lys1158 were modeled as alanines due to poor electron densities for their side chains.

The hPum-Puf5 structure was also solved by MR with the protein from an hPum-hb NRE complex as a search model. The RNA densities were very well defined in both complexes of the asymmetric unit. The RNA register was confirmed with brominated hPum-Puf-5 data. The structure was refined to the resolution limit of the native data (2.32 Å) by using CNS, which lowered the R_{free} to ~29%. At this stage, refinement was continued with REFMAC (Winn et al., 2003), which allows for the anisotropic motion of rigid bodies, described as TLS parameters. Briefly, TLS refinement (using the same test set of reflections used to calculate R_{free} in CNS) was performed with the two proteins and the two RNA chains (in the asymmetric unit) defined as individual rigid units. This lowered the R_{factor} to ~21% and the R_{free} to ~27%. A Ramachandran plot for the refined model shows 91.5% of the residues in the most favored regions, 7.5% in the additionally allowed regions, 1.0% in the generously allowed regions, and none in the disallowed regions. Residues Arg831, Arg840, Arg864, Glu1034, and Lys1053 were modeled as alanines due to poor electron densities for their side chains.

Fluorescence Anisotropy

5′-fluorsceine (5′FL)-labeled $CycB_{direct}$, $CycB_{reverse}$, and Puf5 RNAs were synthesized and deprotected as recommended by manufacturer (Dharmacon Research). Fluorescence emission intensities were collected on a Panvera Beacon 2000 fluorescence polarization system (at 20°C), and the anisotropy values calculated as previously described (Lone et al., 2007). Each reaction sample (total volume of 200 μl) consisted of 1 nM of 5′FL-labeled RNA and increasing concentrations of the protein (from 0.001 nM to 400 nM) in a binding buffer containing 25 mM sodium/potassium phosphate (pH 6.0) and 50 mM NaCl. The samples were left to equilibrate at room temperature for >30 min before the fluorescence anisotropy values were measured. Anisotropy values were referenced against a blank buffer at the beginning of each experiment to account for background correction. Anisotropy values were normalized by first subtracting the anisotropy value with no protein added and then dividing by the maximum anisotropy value for a particular RNA series. Anisotropy values were then plotted versus protein concentration, and the data fitted by nonlinear least-squares regression, by using Origin 7 (OriginLab), to the following quadratic equation:

$$\theta = \left[(K_d + R_o + P_o) - \left\{(K_d + R_o + P_o)^2 - 4R_oP_o\right\}^{1/2}\right]/2R_o$$

where θ is the fraction of RNA bound, R_o is the total concentration of RNA, P_o is the total protein concentration, and K_d is the dissociation constant.

ACCESSION NUMBERS

The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 3BSB (hPum-$CycB_{reverse}$) and 3BSX (hPum-Puf5).

ACKNOWLEDGMENTS

We thank the staff at Advanced Photon Source (beamlines 17-ID and 17-BM) and at Brookhaven National Laboratory (beamlines X4C and X6A) for facilitating X-ray data collection. We thank J. Seetharaman for help with data collection. This work was supported by grant GM062947 from the National Institutes of Health (A.K.A. and R.P.W.). R.P.W. is an Investigator of the Howard Hughes Medical Institute.

Received: November 2, 2007
Revised: December 29, 2007
Accepted: January 2, 2008
Published online: March 6, 2008

REFERENCES

Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., and Kobayashi, S. (1999). Maternal Pumilio acts together with Nanos in germline development in *Drosophila* embryos. Nat. Cell Biol. *1*, 431–437.

Bernstein, D., Hook, B., Hajarnavis, A., Opperman, L., and Wickens, M. (2005). Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. RNA *11*, 447–458.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R., Jiang, W., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Chagnovich, D., and Lehmann, R. (2001). Poly(A)-independent regulation of maternal hunchback translation in the *Drosophila* embryo. Proc. Natl. Acad. Sci. USA *98*, 11359–11364.

Cheong, C.G., and Hall, T.M. (2006). Engineering RNA sequence specificity of Pumilio repeats. Proc. Natl. Acad. Sci. USA *103*, 13635–13639.

Cho, P.F., Gamberi, C., Cho-Park, Y.A., Cho-Park, I.B., Lasko, P., and Sonenberg, N. (2006). Cap-dependent translational inhibition establishes two opposing morphogen gradients in *Drosophila* embryos. Curr. Biol. *16*, 2035–2041.

Crittenden, S.L., Bernstein, D.S., Bachorik, J.L., Thompson, B.E., Gallegos, M., Petcherski, A.G., Moulder, G., Barstead, R., Wickens, M., and Kimble, J. (2002). A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans*. Nature *417*, 660–663.

Dubnau, J., Chiang, A.S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., Smith, P., Buldoc, F., Scott, R., Certa, U., et al. (2003). The staufen/pumilio pathway is involved in *Drosophila* long-term memory. Curr. Biol. *13*, 286–296.

Edwards, T.A., Pyle, S.E., Wharton, R.P., and Aggarwal, A.K. (2001). Structure of Pumilio reveals similarity between RNA and peptide binding motifs. Cell *105*, 281–289.

Gerber, A.P., Herschlag, D., and Brown, P.O. (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol. *2*, E79.

Gerber, A.P., Luschnig, S., Krasnow, M.A., Brown, P.O., and Herschlag, D. (2006). Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *103*, 4487–4492.

Gu, W., Deng, Y., Zenklusen, D., and Singer, R.H. (2004). A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. Genes Dev. *18*, 1452–1465.

Hook, B.A., Goldstrohm, A.C., Seay, D.J., and Wickens, M. (2007). Two yeast PUF proteins negatively regulate a single mRNA. J. Biol. Chem. *282*, 15430–15438.

Jackson, J.S., Jr., Houshmandi, S.S., Lopez Leban, F., and Olivas, W.M. (2004). Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. RNA *10*, 1625–1636.

Jones, T.A., Zou, J.-Y., and Cowan, S.W. (1991). Improved methods for building models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kadyrova, L.Y., Habara, Y., Lee, T.H., and Wharton, R.P. (2007). Translational control of maternal Cyclin B mRNA by Nanos in the *Drosophila* germline. Development *134*, 1519–1527.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *A47*, 110–119.

Lehmann, R., and Nusslein-Volhard, C. (1987). Involvement of the pumilio gene in the transport of an abdominal signal in the *Drosophila* embryo. Nature *329*, 167–170.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. Development *124*, 2463–2476.

Lone, S., Townson, S.A., Uljon, S.N., Johnson, R.E., Brahma, A., Nair, D.T., Prakash, S., Prakash, L., and Aggarwal, A.K. (2007). Human DNA polymerase

kappa encircles DNA: implications for mismatch extension and lesion bypass. Mol. Cell *25*, 601–614.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C., and Read, R.J. (2005). Likelihood-enhanced fast translation functions. Acta Crystallogr. D Biol. Crystallogr. *61*, 458–464.

Menon, K.P., Sanyal, S., Habara, Y., Sanchez, R., Wharton, R.P., Ramaswami, M., and Zinn, K. (2004). The translational repressor Pumilio regulates presynaptic morphology and controls postsynaptic accumulation of translation factor eIF-4E. Neuron *44*, 663–676.

Murata, Y., and Wharton, R.P. (1995). Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in *Drosophila* embryos. Cell *80*, 747–756.

Olivas, W., and Parker, R. (2000). The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. EMBO J. *19*, 6602–6611.

Opperman, L., Hook, B., DeFino, M., Bernstein, D.S., and Wickens, M. (2005). A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. Nat. Struct. Mol. Biol. *12*, 945–951.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Sato, K., Hayashi, Y., Ninomiya, Y., Shigenobu, S., Arita, K., Mukai, M., and Kobayashi, S. (2007). Maternal Nanos represses hid/skl-dependent apoptosis to maintain the germ line in *Drosophila* embryos. Proc. Natl. Acad. Sci. USA *104*, 7455–7460.

Sonoda, J., and Wharton, R.P. (1999). Recruitment of Nanos to hunchback mRNA by Pumilio. Genes Dev. *13*, 2704–2712.

Sonoda, J., and Wharton, R.P. (2001). *Drosophila* Brain Tumor is a translational repressor. Genes Dev. *15*, 762–773.

Tadauchi, T., Matsumoto, K., Herskowitz, I., and Irie, K. (2001). Post-transcriptional regulation through the HO 3′-UTR by Mpt5, a yeast homolog of Pumilio and FBF. EMBO J. *20*, 552–561.

Wang, X., Zamore, P.D., and Hall, T.M. (2001). Crystal structure of a Pumilio homology domain. Mol. Cell *7*, 855–865.

Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M. (2002). Modular recognition of RNA by a human pumilio-homology domain. Cell *110*, 501–512.

Wharton, R.P., and Struhl, G. (1991). RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen nanos. Cell *67*, 955–967.

Wharton, R.P., and Aggarwal, A.K. (2006). mRNA regulation by Puf domain proteins. Sci. STKE *2006*, pe37.

Wharton, R.P., Sonoda, J., Lee, T., Patterson, M., and Murata, Y. (1998). The Pumilio RNA-binding domain is also a translational regulator. Mol. Cell *1*, 863–872.

Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R. (2002). A PUF family portrait: 3′UTR regulation as a way of life. Trends Genet. *18*, 150–157.

Winn, M.D., Murshudov, G.N., and Papiz, M.Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol. *374*, 300–321.

Wreden, C., Verrotti, A.C., Schisa, J.A., Lieberfarb, M.E., and Strickland, S. (1997). Nanos and pumilio establish embryonic polarity in *Drosophila* by promoting posterior deadenylation of hunchback mRNA. Development *124*, 3015–3023.

Ye, B., Petritsch, C., Clark, I.E., Gavis, E.R., Jan, L.Y., and Jan, Y.N. (2004). Nanos and Pumilio are essential for dendrite morphogenesis in *Drosophila* peripheral neurons. Curr. Biol. *14*, 314–321.

Zamore, P.D., Williamson, J.R., and Lehmann, R. (1997). The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. RNA *3*, 1421–1433.

Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., and Wickens, M.P. (1997). A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. Nature *390*, 477–484.